Journal of Experimental Psychology:
Learning, Memory, and Cognition
1989, Vol. 15, No. 3, 527-530

0278-7393/89/$00.75

OBSERVATIONS

Practice, Working Memory, and the ACT* Theory of Skill Acquisition: A Comment on Carlson, Sullivan, and Schneider (1989)

John R. Anderson
Carnegie-Mellon University

The major results reported by Carlson, Sullivan, and Schneider (1989) confirm predictions of the ACT* theory (Anderson, 1983). In particular, ACT* predicts the detrimental effects of the transition to randomized practice because of the need to learn new productions, the complexity effect of gate and judgement type because of more complex production conditions, the effects of practice and its interaction with complexity because of the strengthening mechanisms, and the effects of memory load because of the need to hold information active in working memory so that it can be matched by production conditions.

It is not clear that this note has anything to say that directly contradicts Carlson, Sullivan, and Schneider (1989). I certainly think the experiment and data will prove to be important ones in the study of skill acquisition. Nonetheless, the reader might leave that article with the impression that it claimed to have disconfirmed the ACT* theory (Anderson, 1983). Quite to the contrary, the data in their article are impressive support for the ACT*. The point of this note is to detail the support they offer.

In order to understand exactly what the predictions are for these data, it is necessary to work out in a little detail what the ACT* analysis would be for these experiments. This is what is missing from the Carlson et al. article—which is understandable, given that evaluation of ACT* is not their central purpose.

The first step in working out the ACT* predictions for these data is to devise a set of plausible production rules for performing the task. The assumption of Carlson et al. (1989) is that the complexity effects in their data would be handled in a production system by there being more productions in the case of greater complexity. So let us imagine a production set that would process the information in this way. For simplicity's sake, I will ignore OR gates. In that case, the production rule in Table 1 would have the desired properties and would handle the AND, NAND, BUFFER, and INVERTER gates.[1]

Productions P1 and P2 set up the processing of the two basic types of gates. The basic logic is to assume that the answer will be 1 unless a 0 is encountered. The last step will contain a check for negation of these gates. P3 and P4 scan for 0s, which will reverse the initial hypothesis that the answer will be 1; P5 sets the goal to invert the output if negation is present; P6 and P7 output the value if negation is not present; finally P8 and P9 invert the values. As Carlson et al. note, ACT* predicts that with practice, sequences of these rules will be composed into single rules, one rule for each kind of trial. Let us consider what the compositions of these rules might be. I will only consider compositions that deal with cases of an input of 1, but the rest could be generated analogously:

P1&P4&P7: IF the goal is to predict a gate
and the symbol is BUFFER
and the first input is 1,
THEN output 1.

P1&P4&P5&P9: IF the goal is to predict a gate
and the symbol is BUFFER
and the first input is 1
and negation is present,
THEN output 0.

P2&P4&P4&P7: IF the goal is to predict a gate
and the symbol is AND
and the first input is 1
and the second input is 1,
THEN output 1.

P2&P4&P4&P5&P9: IF the goal is to predict a gate
and the symbol is AND
and the first input is 1
and the second input is 1
and negation is present,
THEN output 0.

With these productions now in place, we can begin to examine some of the major results by Carlson et al. (1989). I will go through the results in the order they are listed in their article.

Preparation of this report was supported by Grant NBS 87-05811 from the National Science Foundation. I would like to thank Kevin Singley for his comments on this article.

Correspondence concerning this article should be addressed to John R. Anderson, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

[1] The rules would have to more complicated to discriminate AND and OR gates, but this would not change any of the basic points.

Table 1
Production Rules Sufficient to Handle AND, NAND, BUFFER, and INVERTER Gates

Production rule	IF	THEN
P 1	the goal is to predict a gate and the symbol is BUFFER	set the answer to 1 and set as subgoals to check that the first input is not 0 output the answer
P 2	the goal is to predict a gate and the symbol is AND	set the answer to 1 and set as subgoals to check that the first input is not 0 check that the second input is not 0 output the answer
P 3	the goal is to check that an input is not 0 and the input is 0	set the answer to 0 and POP the goal
P 4	the goal is to check that an input is not 0 and the input is 1	POP the goal
P 5	the goal is to output the answer and there is negation present	set as a subgoal to invert the answer
P 6	the goal is to output the answer and the answer is 0	output 0 and POP the goal
P 7	the goal is to output the answer and the answer is 1	output 1 and POP the goal
P 8	the goal is to invert the answer and the answer is 0	output 1 and POP the goal
P 9	the goal is to invert the answer and the answer is 1	output 0 and POP the goal

Transition to Randomized Practice

There is a marked increment in time when the various gate types are no longer tested in pure blocks. This is so because the subject does not need tests in the pure block to identify gate type. For instance, rather than P1 & P4 & P7 above, during the buffer block the subject can simply have

IF the first input is 1,
THEN output 1.

This will no longer be useful in the mixed block, and the subject will have to start all over again. The longer reaction times are due to the need to relearn and due to the fact that the productions for the mixed block have larger conditions and will take longer to match. It is an assumption of the ACT* (Anderson, 1983) theory that it takes longer to match productions of greater complexity.

Gate and Judgement Type Effects

There was an effect of gate complexity that lasted throughout the experiment. The initial set of productions, P1 through P9, produced an effect of complexity in that more productions are required to evaluate more complex gates. The claim in Carlson et al. seems to be that this complexity effect should disappear for the composed productions. However, the important observation is that the conditions are larger for the composed productions that deal with the more complex case. Basically, each production combined in the composition leaves some test in the condition of the composed production. This is only reasonable, or else there is no reason to have the production in the original processing. As already noted, it is a basic assumption of the ACT* theory that time to match a production will increase with the number of tests in a production condition. Thus, we predict a complexity effect before or after the composition. Interestingly, this is not a prediction of the SOAR production system theory (Rosenbloom & Newell, 1987).

Carlson et al. report a control experiment in response to this attempt to explain complexity effects in terms of condition size. There they showed that subjects did not take longer to visually discriminate among positive and negative gate types. This is an interesting result in that it shows a difference between a perceptual and semantic discrimination. I am not sure I understand the basis for the difference. This experiment seems to imply that we should not model perceptual discrimination by productions, rather than that we should not model the original task by productions. It seems, in general, that perception is more influenced by issues of stimulus confusability than by stimulus complexity.

One could also incorporate into this model the difference between prediction and verification. Verification would require extra productions or conditions to check match between the predicted output and the actual output.

Practice

As noted by Carlson et al., the basic shape of the practice curves is consistent with ACT* among other theories. Basically, the power function is a result of ACT*'s strengthening mechanism and not the composition process (Anderson, 1982). Much is made of the fact that the gate complexity effects remain even after more than a 1,000 trials of practice per gate. However, the conditions do converge with practice. ACT* does predict that as the productions become strengthened, the time to match the production will go to zero, leaving the remaining time a constant associated with perceptual and motor processes. The critical question is whether the learning curves are converging to this perceptual–motor asymptote as rapidly as ACT* predicts. Carlson et al. makes much of the fact that the curves are still significantly apart after more than 1,000 trials per gate. However, power functions converge very slowly. As Pirolli and Anderson (1985) develop them, the ACT* predictions are approximated by a power function of the form

$$I + A\,P^{-b},$$

where I is intercept, A is a factor depending on condition, P is trials of practice, and b is the learning rate. Such a power function starts at $I + A$ on Trial 1 and converges to I. Thus, the curves differ initially because of the condition-specific A, but all converge to the same time.

I fit functions of this form as best I could to the data in Figure 4 from Carlson et al. I tried constraining all possible (2^3) combinations of the parameters I, A, and b to be constant across conditions. Table 2 reports the results in terms of parameter values and percentage of variance accounted for by various parameter combinations. Although there is naturally improvement of fit with every free parameter, the data clearly indicate that the best compromise between numbers of free parameters and fit is obtained when just the A parameter is free to vary among conditions as the ACT* theory would predict (Model 5 of Table 2). Letting the b and I parameters to vary allows for little further improvement in fit. The Carlson et al. model seems to be one in which both the I and the A parameters are allowed to vary among conditions because they assume different asymptotes (I) but also that there is some convergence (hence the need for different A). Under the ACT* model, 97.7% of the variance is fit (Model 5 in Table 2), and only an additional 0.2% is gained by adding free intercept parameters for the Carlson et al. model (Model 3 in Table 2). In addition, the intercept parameters are not well behaved in Model 3, with the lowest value (15 ms) found in the most difficult condition.

Figure 1 displays the data and the best fitting curves under Model 5. Clearly, the conditions are not converging more slowly than the ACT* theory would predict. Although the conclusions from this exercise must be regarded as tentative, because I had access only to Figure 4 from Carlson et al. and not the raw data,[2] this exercise does provide an interesting object lesson concerning the properties of power functions. To get these curves to within 10 ms of each other would require 7,200,000 trials!

Memory Load Task

The matching of ACT* productions is influenced by the level of activation of the data structures to which the condition patterns are being matched. The lower the level of activation, the slower the pattern matching. Also as noted in Anderson (1983), manipulations of level of data activation will interact with manipulations of pattern complexity in a superadditive way. Note that the relevant factor is not how much information is being maintained in working memory but rather the level of activation of just that information which is being used to match a condition. Thus, we would not predict an effect of extra working memory load per se. This is consistent with the Carlson et al. report.

Generally, the relevent data with which to match the production condition are held on the screen in front of the subject and should be maintained in a high level of activation. The one exception to this is the access condition, where the information about the input to the gates was maintained internally. The ACT* model for maintaining memory load in this task is basically the model for maintaining a memory set in the Sternberg task (Anderson, 1983; Jones & Anderson, 1987). In this model, the subject has to maintain the elements of the memory set (in this case variable assignments) in an active state so that they can be matched by production conditions. This is achieved by having a node in memory representing the set which is a focus of attention and hence kept active. Activation spreads from it to the members of the set. This basically creates a fan effect in which the amount of

[2] This is so only because of time pressures in preparing the reply. Carlson has agreed to send me the raw data.

Table 2
Parameter Values and Percentage (%) Variance Accounted for Under Various Sets of Constrains on the Parameters I, A, b

	Predict/Pos			Predict/Neg			Verify/Pos			Verify/Neg			
Model	I	A	b	I	A	b	I	A	b	I	A	b	% variance
1	134	4,204	.31	0	4,484	.26	294	6,384	.40	0	6,477	.30	98.0
2	$\underline{87}$	4,077	.29	$\underline{87}$	4,628	.28	$\underline{87}$	5,316	.32	$\underline{87}$	6,742	.32	97.9
3	122	4,165	$\underline{.31}$	159	4,822	$\underline{.31}$	68	5,201	$\underline{.31}$	15	6,539	$\underline{.31}$	97.9
4	393	$\underline{6,275}$	.47	393	$\underline{6,275}$	.42	283	$\underline{6,275}$	.39	0	$\underline{6,275}$	.30	97.7
5	$\underline{189}$	4,533	$\underline{.34}$	$\underline{189}$	5,549	$\underline{.34}$	$\underline{189}$	5,425	$\underline{.34}$	$\underline{189}$	6,679	$\underline{.34}$	97.7
6	$\underline{371}$	$\underline{6,954}$	.48	$\underline{371}$	$\underline{6,954}$	.43	$\underline{371}$	$\underline{6,954}$	.44	$\underline{371}$	$\underline{6,954}$	.39	96.9
7	0	$\underline{5,314}$	$\underline{.32}$	143	$\underline{5,314}$	$\underline{.32}$	114	$\underline{5,314}$	$\underline{.32}$	280	$\underline{5,314}$	$\underline{.32}$	95.5
8	$\underline{104}$	$\underline{5,220}$	$\underline{.31}$	$\underline{104}$	$\underline{5,220}$	$\underline{.31}$	$\underline{104}$	$\underline{5,220}$	$\underline{.31}$	$\underline{104}$	$\underline{5,220}$	$\underline{.31}$	86.2

Note. Pos = positive; Neg = negative; I = intercept; A = factor depending on condition; b = learning rate. Underlined parameter values were constrainted to have the same values in the model fit.

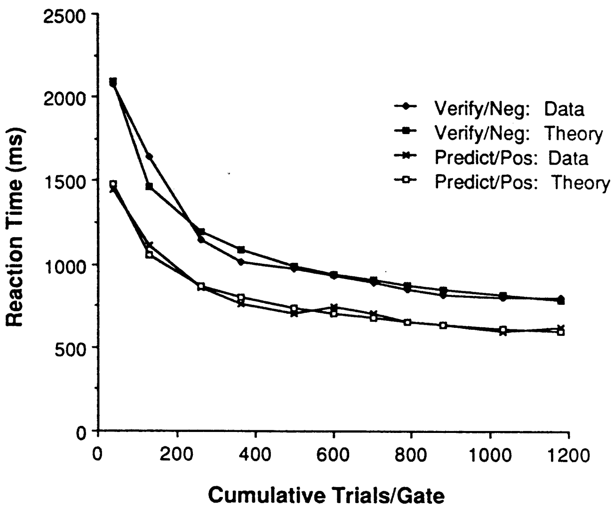

Figure 1. Predicted and empirical curves for the Predict/Pos and Verify/Neg. (The empirical values were extrapolated from Figure 4 of Carlson et al., 1989.) The abscissa values are 200 trials less than ones in the Carlson et al. to discount the first 144 pure trials, plus the fact that the abscissa trials in Figure 4 appear to be last trial in a block rather than average trial. (Pos = positive; Neg = negative.)

activation reaching an element is diminished in proportion to the number of elements being kept active. In particular, the activation level of the elements that will be used to match the production conditions is diminished. Hence, ACT* predicts the longer times with greater set size in the access condition.

Summary

I was frankly surprised at how well the Carlson et al. (1989) experiment corresponded to the ACT* theory. The motivation of this comment was to emphasize this confirmation rather than to take issue with Carlson et al. Given that their experiment was manifestly not designed with confirmation of the theory in mind, I think this is a remarkable result. In particular, the theory predicts or is consistent with the four results stressed by Carlson et al.—the effect of the switch from blocked to randomized practice, the effect of task complexity, the interaction between complexity with practice, and the effect of memory load.

It might seem peculiar that an experiment that was cast as relevant to the composition mechanisms of ACT* should prove more relevant to the strengthening mechanism. This raises the question of what experimental tests have been performed on composition. The best experiment I know of is that of McKendree and Anderson (1987). We had subjects evaluate combinations of LISP functions so that each individual function was evaluated equally often but the combinations were encountered with different frequencies. There we found evidence that subjects evaluated the more frequent combinations more rapidly, indicating that they had composed the basic functions into combinations and were differentially strengthening these combinations.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge. MA: Harvard University Press.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 15, 517–526.

Jones, W. P., & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. *Journal of Experimental Psychology: General, 116,* 137–153.

McKendree, J. E., & Anderson, J. R. (1987). Frequency and practice effects on the composition of knowledge in LISP evaluation. In J. M. Carroll (Ed.), *Cognitive aspects of human–computer interaction* (pp. 236–259). Cambridge, MA: MIT Press.

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 136–153.

Rosenbloom, P., & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 221–286). Cambridge, MA: MIT Press.

Received September 20, 1988
Revision received October 11, 1988
Accepted October 17, 1988 ■